A METHOD TO REMOVE SIZE BIAS IN SUB-CORTICAL STRUCTURE SEGMENTATION

Mythri V, Alphin J Thottupattu*, Naren Akash R J, Jayanthi Sivaswamy*

Centre for Visual Information Technology, IIIT Hyderabad, India

Segmentation and analysis of sub-cortical structures is of interest in diagnosing some neurological diseases. Segmentation is a challenging task because of brain tissue ambiguity and data scarcity. Deep learning (DL) solutions are widely used for this purpose by considering the problem as a semantic segmentation of the brain. In general, DL approaches exhibit a bias towards larger structures when training is done on the whole brain. We propose a method to address this problem wherein a pre-training step is used to learn tissue characteristics and a rough ROI extraction step aids focusing on local context. We use a Residual U-net for demonstrating the proposed method. Experiments on the IBSR and MICCAI datasets show that our proposed solution leads to an improvement in segmentation performance in general with medium and small size structures benefiting significantly. The performance with the proposed method is also marginally better than a more complex, state of art sub-cortical structure segmentation method. A strength of the proposed method is that it can also be applied as a modification to any existing segmentation solution.

1. INTRODUCTION

Sub-cortical (unlike cortical) structures are grey matter structures embedded in white matter which vary widely in size. This poses a challenge for their segmentation. With the popularity of deep learning, deep neural networks have been explored extensively for segmentation including the sub-cortical structures [1]. The U-net [2] has proven to be a resilient blueprint for segmentation tasks in general since its encoder-decoder design enables learning from smaller medical image datasets.

Many variants of the U-net have been introduced for sub-cortical structure segmentation [3, 4, 5] with reasonable improvements. A label refinement strategy with two concatenated U-nets is used in [5] and [3] introduced competitive dense blocks and multi-slice information aggregation. ψ-net [4] uses a densely convolutional LSTM module for selecting and enhancing features. Wang et al.[6] proposed a Residual Attention Network which uses self-attention to compute 3D attention maps. Hu et al.[7] proposed a compact module called squeeze and excitation block to compute channel-wise attention.

All existing methods are agnostic to the size variation among the structures to be segmented and pass the whole volume or fixed size patches. This creates a class imbalance while training. A performance bias can be observed towards the segmentation results for bigger structures in most of the proposed solutions. Possible reasons include i) inconsistency in annotations of large vs. small structures across subjects, as the former are easier to annotate compared to the latter, ii) insufficient image resolution leading to more inaccurate boundaries for smaller structures and iii) learning both large and small structures from many (for the former) versus few (for the latter) voxels is inefficient. While the first two issues cannot be resolved easily, the last one can be resolved by taking a different approach. Towards this end, in this paper, we propose to make the network learn the structures from its local rather than global context such as the whole brain image. This should aid the network in focusing on each structure separately irrespective of their size, giving an edge to smaller structures. Additionally, we propose a pre-training step to learn tissue types to aid in discriminating between tissue types, as the local context information alone can be inadequate for efficient segmentation.

2. METHOD

The proposed method focuses on narrowing the disparity in performance between structures of different sizes. To this end, we propose a 2 phase segmentation framework. The phases are: 1) Grey matter (GM), White matter (WM), Cerebrospinal fluid (CSF) segmentation for pre-training. 2) Sub-cortical segmentation with atlas-guided ROI extraction.

The objective of phase 1 is to help the network learn to discriminate between the main tissue types in the brain, (GM, WM and CSF), which in turn should aid sub-cortical structure segmentation since these structures are GM embedded in WM deep inside the brain. In this pre-training phase, 3D volume is fed to the network with GM, WM and CSF labels, computed using FSL[8] as ground truth.

In the next phase, weights of previous phase are loaded as initial weights and a rough ROI patch is extracted for each of the sub-cortical structures and fed to the Network with the structure labels as ground truth. The output after this phase of training is the probability map for each structure. The prob-

*These authors contributed equally to this work

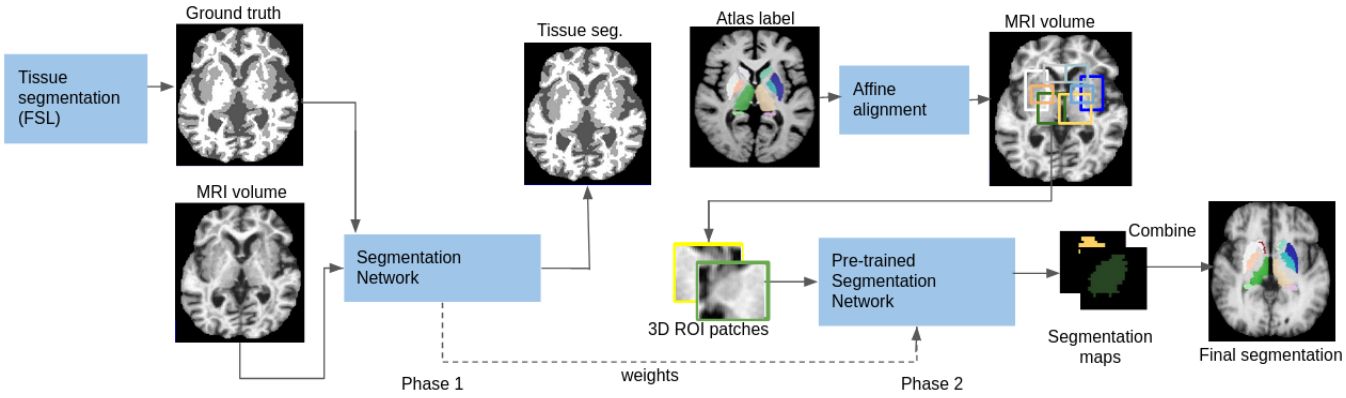

Fig. 1. Proposed 2-phase training framework for sub-cortical structure segmentation. Phase 1: Pre-training , Phase 2: Structure segmentation from the 3D structure ROI.

ability maps are then combined, with maximum probability at each pixel to get the final segmentation map. The motivation behind this strategy is to counteract the variability in sizes of the structures, it ensures that smaller structures get equal priority while training compared to methods that provide sub-cortical region/whole volume as input. In our experiments, the rough ROIs are extracted by an affine alignment of the atlas to the MRI volume. A relaxation step (expanding ROI size by a few pixels) was done for each 3D ROI patch to compensate for imprecise alignment and to prevent under-segmentation. The two phases of training that has been proposed network are illustrated in Fig. 1.

2.1. Implementation Details

We employ 3D U-Net[9] based architecture with Residual Blocks with 3 encoder-decoder levels. Each Residual block consists of two 3D convolutional layers followed by Instance Normalization and ReLU-activation. After each block of the encoder, downsampling is performed using max pooling layer and upsampling is done using transpose convolution layer. Adam optimizer is used with initial learning rate 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and weight decay of 0.001. Categorical cross entropy is used as the loss function for phase 1. A sum of unweighted soft dice loss and categorical cross entropy loss is utilized to train the model in phase 2. The code was implemented using PyTorch and training was done on NVIDIA GTX 2080 with 11GB RAM. A fixed atlas [10] was used for extracting ROIs in training and testing phases.Implementation is available at `https://github.com/mythri-venkat/subcortical_segmentation`.

3. EXPERIMENTS

Two publicly available datasets were used to perform experiments: IBSR [11] and MICCAI [12] datasets with 18 and 35 volumes, respectively. The voxel size in IBSR is variable, namely, $0.93 \times 0.93 \times 1.5$ or $1 \times 1 \times 1.5$ whereas in MICCAI dataset it is uniformly $1 \times 1 \times 1.25$. Both these datasets provide labels for both cortical and sub-cortical structures. Volume is cropped to sub-cortical region (80x80x80 voxels). A 6-fold validation was done on IBSR dataset while a 5-fold validation was done for MICCAI and the results were averaged across all the folds for reporting. In order to understand the size induced performance variations, analysis was done with a sizewise ordering of structures, i.e. structures are divided into 4 classes: very small, medium, large and very large. In terms of volume, these are approximately 4%, 20%, 50% of the average size of the largest structure in the 'very large' category (namely the thalamus)

Quantitative evaluation of segmentation with respect to ground truth was done using the Dice Similarity Coefficient (Dice) [13]. Dice helps assess the overlap between the predicted segmentation (A) result and the ground truth (B). It is defined as $Dice(A, B) = \frac{2\times|A\cap B|}{|A|+|B|}$. Dice value lies between 0 and 1, where former constitutes no overlap and latter represents complete overlap with ground truth.

4. RESULTS

The performance of a Residual U-net with the proposed 2-phase training approach was assessed by comparing against plain Residual U-net trained with volume cropped to sub-cortical region. The obtained results are presented in Table 1. Dice values are given for the structures at an individual as well as at a group (sizewise) level. In order to understand the contribution from each phase of training, ablation studies were done. The three variants considered were: Segmenting after pre-training (PT), after ROI extraction (ROI), and with the proposed 2-phase training. These were evaluated on the IBSR and MICCAI datasets and results are reported in Table 1. A performance degradation of 22% is observed in Dice value for the very small structure group relative to very large class

	IBSR Dataset					MICCAI Dataset				
	Dice (Mean ± STD)		Av. Improvement(%)			Dice (Mean ± STD)		Av. Improvement(%)		
Left & Right	**Res. U-net (baseline)**	**Proposed**	**PT& ROI**	**PT**	**ROI**	**Res. U-net (baseline)**	**Proposed**	**PT& ROI**	**PT**	**ROI**
1. Very Small										
Accumbens	0.697 ± 0.049	0.735 ± 0.039	5.55	2.1	3.3	0.75 ± 0.084	0.783 ± 0.044	4.8	1.8	4.05
2. Medium										
Amygdala	0.736 ± 0.055	0.766 ± 0.048	3.7	1.325	3.4	0.776 ± 0.046	0.816 ± 0.025	4.35	3.5	4.225
Pallidum	0.804 ± 0.025	0.831 ± 0.021				0.845 ± 0.076	0.875 ± 0.025			
3. Large										
Caudate	0.868± 0.018	0.876± 0.02	1.3	0.533	0.617	0.866 ± 0.10	0.877 ± 0.050	1.56	0.5	1.45
Hippocampus	0.809 ± 0.021	0.824 ± 0.023				0.846 ± 0.028	0.863 ± 0.021			
Putamen	0.884 ± 0.015	0.895 ± 0.012				0.893 ± 0.061	0.905 ± 0.035			
4. Very Large										
Thalamus	0.893 ± 0.008	0.904 ± 0.008	1.3	0.75	0.55	0.901± 0.035	0.914 ± 0.021	1.4	0.95	1.35
Av. 2-4	0.832 ± 0.023	0.849 ± 0.022				0.855 ± 0.058	0.875 ± 0.029			
Av. full	0.814 ± 0.027	0.834 ± 0.024				0.839 ± 0.061	0.862 ± 0.031			

Table 1. Performance analysis of the proposed method. Dice scores averaged over 6 or 5 folds are listed for a baseline Res. U-net and its variants: Trained with Proposed method, only with pre-training (PT) and only with ROI training (ROI).

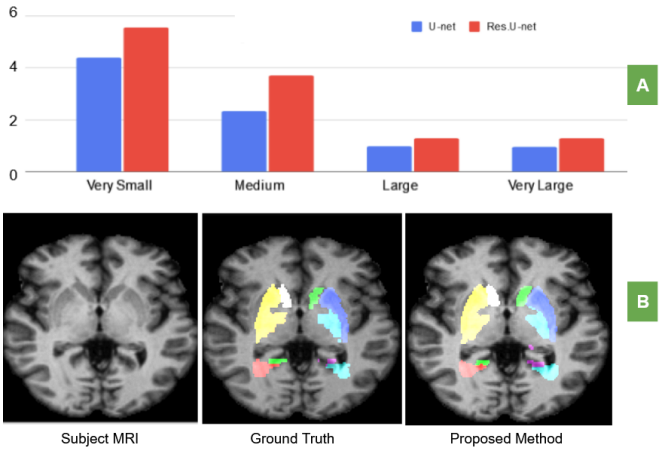

Fig. 2. For IBSR dataset A) Improvement in Dice score (%) of structures of different sizes for proposed method.Base line is a plain U-Net and Res. U-Net B) Qualitative results for proposed method (Res. U-Net)

group for baseline Res. U-net. The state of the art (SOTA) ψ-net also reports a 17.6% deterioration[4]. This underscores the need for developing a strategy to overcome size-specific performance variations. The Dice value for smaller structures in both the datasets show improvements with the proposed modifications in training compared to the larger ones. Another observation from Table 1 is that the improvement with the proposed method is significant from both phase 1 and 2 for smaller structures compared to larger ones. Results for MICCAI dataset appear to be more influenced by pre-training and ROI-based training compared to IBSR. This is possibly due to the difference in data quality. In MICCAI (relative to the IBSR images), the voxel resolution in the axial direction is better; the contrast between tissues and the quality of the image is also superior. Thus, the proposed method appears to boost segmentation performance in general with the quantum of boost being affected by the data quality/resolution which is a fundamental issue for segmentation.

Our proposed training strategy was applied to a vanilla U-net to determine the generalizability of the proposal. The improvement in Dice scores obtained for the different groups of structures is presented as a bar graph in Fig. 2(A). The trends of improvement obtained with Res U-net is also consistently replicated for the U-net, even though the quantum of improvement is less for U-net, which is to be expected. Sample outputs of the proposed method for IBSR dataset are shown in Fig. 2(B) for visual comparison and it is observed that the labels obtained with the proposed method are very close to the ground truth with smooth boundaries.

	Dice
Freesurfer	0.74 ± 0.11
FIRST	0.81 ± 0.08
U-net	0.80 ± 0.03
Res. U-net	0.81 ± 0.03
ψ-net	0.82 ±0.03
Proposed Method	0.83 ± 0.02

Table 2. Performance comparison with standard (available in toolboxes) and state of the art solution for sub-cortical structure segmentation on IBSR dataset.

Finally, a comparison is made with baseline segmentation